MODIS POSTLAUNCH DATA PROCESSING SCENARIOS

D. Han
NASA/Goddard Space Flight Center
Greenbelt, MD

P. Ardanuy, W. Gregg, D. Hoyt, A. McKay, and G. Riggs
Research and Data Systems Corporation
Greenbelt, MD

M. Andrews and T. Wolford
General Sciences Corporation
Laurel, MD

December 1989

FOREWORD

This document is one of a set of four completed in late 1989. Specifically, these are the:

- MODIS Core Data Product and Algorithm Report
- MODIS Utility and Support Algorithm Requirements
- MODIS Core Data Product Processing Scenarios
- MODIS Post-Launch Data Product Processing Scenarios

In the present form, all reviewers' comments have been addressed. In addition, the Core Data Product Processing Scenarios document contains a revised set of core products as identified during the January 31 to February 2, 1990 MODIS Science Team Meeting. Of course, the concept of MODIS data product algorithms, their requirements, and operations concepts continues to evolve. Therefore, these documents should be taken as a statement of progress, rather than a definitive treatment of the subjects. As appropriate, revised documents will be distributed to reflect a more comprehensive and quantitative understanding of the MODIS data processing functional, performance, and operational requirements.

EXECUTIVE SUMMARY

This document presents a concept for processing the Moderate Resolution Imaging Spectrometer (MODIS) data and the generation of MODIS data products in the post-launch period. This document discusses some initial concepts regarding the selection, development, and refinement of algorithms among MODIS team members for the production of oceanic, terrestrial, and atmospheric-sciences data products after launch, particularly Level-2 and higher, to meet the Earth Observing System (EOS) mission objectives. Additional motivations include the desire to understand the growth of the MODIS processing requirements during the post-launch period beyond those of the initial core data product set. We also wish to illuminate some concepts regarding the calibration and validation of the MODIS measurements, and the anticipated reprocessing of some of the MODIS data sets.

Six issues are considered in this scenario for MODIS data processing: (1) the conversion of science-team-member proposed research and development (R&D) products into new standard products; (2) the identification of new R&D and standard data products by the general science community, including the interdisciplinary investigators; (3) the development of an accurate estimate growth in the processing and storage requirements of the data system; (4) the maintenance of the calibration of the MODIS measurements; (5) the reprocessing requirements; and (6) the initial post-launch and ongoing validation of MODIS data products (which includes the role and requirements of field experiments). As a result of this analysis, the following general requirements may be stated:

a. EOSDIS must be sized to accommodate growth in the processing requirements of MODIS over five-year periods.

b. The data management system within EOSDIS must be designed to accommodate an increasing number of data products, and possibly an increasing complexity in ancillary data requirements and the corresponding external and internal interfaces.

c. Policies and automated procedures must be developed to efficiently manage the reprocessing of MODIS data products.

d. Adequate capabilities for the transportation and visualization of data sets must be built into EOSDIS to permit the science team members to optimally validate their data products.

TABLE OF CONTENTS

1 INTRODUCTION

This document presents a concept for processing the Moderate Resolution Imaging Spectrometer (MODIS) data and the generation of MODIS data products in the post-launch period. One purpose of this document is to put forward for discussion some initial concepts regarding the selection, development, and refinement of algorithms among MODIS team members for the production of data products after launch, particularly Level-2 and higher, to meet the Earth Observing System (EOS) mission objectives. Additional motivations include the desire to understand the growth of the MODIS processing requirements during the post-launch period beyond those of the initial core data product set. We also wish to illuminate some concepts regarding the calibration and validation of the MODIS measurements, and the anticipated reprocessing of some of the MODIS data sets.

A previous release, entitled the "MODIS Core Data Product Processing Scenarios," reviewed our present state of knowledge concerning the generation of derived MODIS products near launch. Here, we consider the evolution of the processing scenario over a multi-year (five or more years) period following launch. It is possible to identify six issues which must be considered in a scenario for MODIS data processing in the post-launch period. These are:

1. The conversion of science-team member proposed research and development (R&D) products (to be developed within the Team Member Computing Facilities (TMCFs)) into new standard products (to be generated routinely at the Central Data Handling Facility (CDHF)).

2. The identification of new R&D and standard data products by the general science community, including the interdisciplinary investigators.

3. The specification of an accurate estimate growth in the processing and storage requirements of the processing system over a five-year period.

4. The maintenance of the calibration of the MODIS measurements.

5. The reprocessing requirements over a five-year period.

6. The initial post-launch and ongoing validation of MODIS data products, including the role and requirements of field experiments.

This document is partitioned into six sections. In Sections 2, 3, and 4, post-launch ocean, terrestrial, and atmospheric sciences data processing scenarios are discussed. Issues concerning the integration of MODIS data products are discussed in Section 5. In Section 6, a scenario for the maintenance of the MODIS calibration is discussed. Finally, in Section 7, we consider scenarios related to the reprocessing of MODIS data products.

At the MODIS launch date, core data product algorithms will be in place at the CDHF, methods for their archival and distribution defined, and mechanisms for obtaining ancillary data sets prescribed and in place. Primary post-launch data processing scenario issues involve:

a. continual sensor calibration
b. core data product validation
c. development of Science Team Member R&D data products
d. use of these data products and further research developments by members of the scientific community other than Science Team Members.

This post-launch processing scenario is dependent upon development of algorithms for production of identified core data products, developed from simulated MODIS sensor data and other sensors in the pre-launch execution phases, to be operational at both the TMCF and CDHF. The CDHF will produce these identified core data products routinely in post-launch processing.

In the post-launch period the TMCFs will be supplied with Level-1 calibrated radiances (top of the atmosphere radiance) and Level-2 surface-leaving radiances (calibrated radiances corrected for atmospheric effects) from the CDHF and will be developing and producing Level 2, 3 and 4 (R&D)[1] data

---

[1]R&D products, unlike the core data products, will not be developed and validated by launch. In many cases, the additional observing capabilities of the MODIS instrument will be required to yield the retrieval algorithms.

products. These R&D products will have to pass through some acceptance procedure before being accepted as core data products to be produced at the CDHF.

Initially, input on the development of additional standard products beyond the initial core set will be from the MODIS team members only. Input on product development and the R&D products themselves may be expected to come from the scientific community at large in later post-launch phases.

2 POST-LAUNCH OCEAN SCIENCES DATA PROCESSING SCENARIO

2.1 Core Data Product Validation

For at least the first six months after launch, intensive core product validation efforts should take place to ensure the quality and validity of the core products (not only for the ocean products, but for those of the other disciplines as well). These validation efforts should be performed by the Science Team Member(s) responsible for the core product. Thus adequate communication links between the CDHF and the TMCF should be established and operational for the post-launch validation procedures to be effective. In addition, in-situ ship observations and drifting buoy data used by one or more of the Science Team Members should be made available to all members through the CDHF in order to facilitate the validation efforts. This dissemination will require Science Team Members to submit their in-situ data for archival.

Validation efforts may very likely continue after the first six months following launch to ensure product accuracy, but the frequency and duration will likely be product dependent. Checks may be made on an annual basis to confirm data quality, or an in-situ network of buoys and ships of opportunity may be utilized in an ongoing, continuous validation exercise. These later checks will probably be performed by the Science Team Members; however, they may be performed within EOSDIS as a routine part of the processing scenario.

2.2 Research and Development Products

R&D products for oceans are listed in Table 1. In nearly all cases, these products are derived from Level-2 ocean core data products, and so do not impact the core data product processing scenario as delivered previously. Rather, they are additional candidate standard products beyond the core data products, generated after most of the core products have been produced.

Table 1. Ocean Research and Development Products for MODIS	
PRODUCT	SCIENCE TEAM MEMBER
Chlorophyll Fluorescence Yield	Dr. Mark Abbott
Australian Data Products	Dr. John Parslow
Mid-Atlantic Bight Data Products (chlorophyll concentrations, phycoerythrin, water-leaving radiance, etc. for this region)	Drs. Frank Hoge and Wayne Esaias
Total Suspended Materials (Case I and II waters)	Drs. Dennis Clark and Kendall Carder
Case II Pigment Concentration	Dr. Kendall Carder
Regional Basins Primary Production (Case I and II waters)	Drs. Mark Abbott and Wayne Esaias
Gelbstoff Concentrations	Dr. Kendall Carder
Detritus Concentration	Dr. Kendall Carder
Sea Surface Temperature Quality Assessment Field	Dr. Otis Brown
Objectively Analyzed Sea Surface Temperatures	Dr. Otis Brown

The exception to this scenario is with respect to Carder's R&D products (chlorophyll, suspended matter, gelbstoff, and detritus). Although Case II chlorophyll is listed as a core product, its inclusion in the R&D products reflects an expectation of substantial improvement given information on the distributions and abundances of suspended matter, gelbstoff, and detritus. It also reflects an expectation that the ocean core products algorithms will be improved significantly prior to the EOS launch. It is reasonable to assume that the years of experience gained from SeaWIFS[2] will allow improvement of the existing proposed algorithms.

These R&D products may then have substantial impact on the Case II chlorophyll products, and may also affect the Case II primary production core data products, since they depend upon knowledge of chlorophyll. These products, as well as the other R&D products, are discussed in detail below.

2.2.1 *Chlorophyll Fluorescence Yield (Dr. Mark Abbott)*

Dr. Abbott plans to obtain estimates of chlorophyll fluorescence yield and its temporal and spatial variations with MODIS-T. Artificially stimulated (with a constant blue light source or a xenon strobe light) chlorophyll fluorescence has been used to gauge chlorophyll concentrations for the last two decades. Controversy has surrounded the technique since its inception. Several environmental and physiological parameters (nutrient stress, light adaptation, and physiological state) can influence the florescence yield. It is also strongly species-dependent with a large diurnal component. It is remarkable that this technique has become the standard in spite of these disadvantages. This speaks to the weaknesses of ship-derived observations.

Sunlight-stimulated natural fluorescence has the same functional dependencies when used to measure chlorophyll concentrations. However, natural in-situ chlorophyll fluorescence may be a useful indicator of photosynthesis. Consequently, MODIS-T would permit rapid, cost-effective surveys to be made. Determination of chlorophyll fluorescence from remote measurements must be

---

[2]The launch of the Sea-viewing, Wide-Field-of-View Sensor (SeaWIFS) is anticipated in 1992; the accumulation of a five-year data set is a mission goal.

supported by routine field collection of calibration samples. This capability does not exist using conventional radiocarbon uptake experiments.

Pre-launch data from SeaWIFS may improve this algorithm. This may be one R&D algorithm which can be made a routine MODIS-T product before its launch. Algorithm development will occur at the TMCF, and will be sent to EOSDIS for incorporation into the post-launch processing stream.

2.2.2 *Australian Data Products (Dr. John Parslow)*

Dr. Parslow proposes to address questions related to sensor calibration, atmospheric correction, and ocean color-constituent relationships in Australian waters. To do this he proposes to undertake a field program to measure ocean color and optical properties in Australian coastal and offshore waters. He will use an 8-band Australian Ocean Color Scanner (OCS), and the Australians' expertise in satellite image processing using in-house developed algorithms. These flights and algorithms can provide ground truth from which to calibrate the MODIS-T sensors immediately after launch.

Dr. Parslow will study the relationship between ocean color and the dissolved and particulate constituents in Australian waters. Using MODIS-T observations he will develop algorithms that will estimate the constituent concentration (organic and inorganic). This work will use laboratory and field measurements of absorption and backscattering coefficients associated with these constituents (i.e., phytoplankton, particulate detritus, gelbstoff, and suspended sediments).

At post-launch, Dr. Parslow proposes to act as a center for the collection, archiving, and processing of MODIS-T data collected over Australian waters. These products will have a practical application to fisheries problems in Australian and more theoretical applications to investigating the carbon flux in the sub-tropical convergence near Australia.

The first problem facing Dr. Parslow is how to obtain the MODIS-T data in Australia. There are no high quality data lines to Australia from the United States. Fiber optical cables are now being laid across the Pacific Ocean, but they are not yet operational. An alternative is to directly downlink data from the EOS satellite as it passes over Australia. Presently this capability is not funded for EOS. Another possibility is to directly receive data from Tracking and Data Relay Satellite System when it is available. The direct downlinking of data currently appears as the most feasible alternative.

Level-2 data products produced will be supplied to the CDHF on a mutually agreed on media after acquisition and processing. Regional products will then be made available to other scientific personnel through EOSDIS.

Dr. Parslow will have field observations from buoys and an Ocean Color Scanner to provide ground truth. These products will require digitizing, resampling, and Earth location before use.

2.2.3 *Mid-Atlantic Bight Data Products (Drs. Frank Hoge and Wayne Esaias)*

Dr. Hoge's stated objectives are to improve estimates of biomass, primary productivity, and to refine phytoplankton pigment measurements. The goal is to obtain a more quantitative carbon and nitrogen global cycling estimates during his 5-year project.

To achieve this will require improving the chlorophyll algorithm by including measurements of phycoerythrin and phycocyanin in addition to chlorophyll a pigments, using active-passive correlative spectroscopy techniques. Further improvements would result from quantifying algorithms which would remotely measure different phytoplankton and zooplankton species using water-leaving radiance values. Remote measurements would permit determining their temporal and spatial variations. After these algorithms are developed, they will apply them to routine MODIS processing.

Validation efforts would continue after the algorithm development phase. They propose to validate the algorithms using NASA's Airborne Oceanographic Lidar (AOL), Advanced Very High Resolution Radiometer (AVHRR) data, and surface vessels. The surface vessels will provide species identification and phyto-plankton pigment concentrations. Perhaps data from Dr. Clark's buoys will be used for validation.

These projects will commence after the EOS launch. Preliminary work will include validation underflights during MODIS passes. Ships will provide surface ground truth measurements.

The bulk of the work could be done a Goddard because of the proximity of Wallops Island to Washington, D.C. If the R&D is done at the Science Team Member's TMCF, then a high quality data communication line to transfer MODIS data will be required. The TMCF will need the capability to produce the regional product. It is not clear whether the Mid-Atlantic Bight product will be produced routinely or occasionally.

2.2.4 *Total Suspended Materials (Dr. Dennis Clark)*

Dr. Clark suggests an elegant but conceptually simple experiment to validate MODIS products. Dr. Clark has proposed to set eight buoys--three at NOAA buoy sites and five drifting in the world's oceans. These buoys will telemeter data back to EOSDIS through GOES. His approach is to install the most sophisticated, currently available optical instrumentation in the buoys to produce high resolution in-water and above-water spectra.

Dr. Clark's strategy is to use optoacoustic spectroscopy (OAS). OAS accurately measures absorption coefficients for specific pigments. OAS works by irradiating particles with a modulated light source. When the particles absorb the light, they thermally expand and contract producing acoustic waves. These acoustic waves are detected with a hydrophone. This technique is insensitive to scattering. Consequently, it is suitable for studying extremely low in-situ concentrations.

A differential OAS can measure phytoplankton absorption in discrete samples at various wavelengths. This permits accurate measurements of the organic and total particulate spectral absorption and scattering properties. Accurate measurements made with these instruments should permit models to be developed (and validated) which will determine the concentration of suspended particulate matter.

Data produced from these buoys are to be delivered (via telemetry through GOES) to EOSDIS for archival and subsequent distribution. The data will be stored as collected, with Earth-location, time, and other ancillary data. This project will directly produce no algorithms or maps for MODIS processing, but rather serve as an in-situ data source for validation of core data products and for the development of new products. Thus, although this project produces no MODIS data products per se, it is essential for the refinement of existing core data products and the development of new scientific techniques. Its direct impact on the post-launch processing scenario, however, is limited, requiring only the archiving of data sets for subsequent distribution.

2.2.5 *Case II Pigment Concentrations*

2.2.6 *Regional Basins Primary Production (Drs. Mark Abbott and Wayne Esaias)*

Dr. Abbott proposes to develop models which can determine the "new" fraction of the total primary production. This represents the carbon amount which may be removed from the ocean's euphotic zone. The downward carbon flux from the euphotic zone is roughly equal to the new production. The new production, in turn, is supported by an upward nitrate flux into the euphotic zone from deeper water. These fluxes vary in both time and space.

Existing data are based on scattered ship observations which are heavily biased toward Northern Hemisphere coastal waters. The Joint Global Ocean Flux Study (JGOFS) data should help to reduce this deficiency. JGOFS should also provide in-situ observations for model development and validation. Additional non-EOS data required for data analysis and modeling are:

a. SeaWIFS
b. Coastal Zone Color Scanner (CZCS)
c. Ocean Color Temperature Sensor (OCTS)
d. SCAT
e. ALT
f. AMSR
g. Advanced Medium Resolution Imaging Radiometer (AMRIR)

This new production couples the atmospheric and oceanic carbon cycles. The variables in phytoplankton biomass and productivity fields are largely driven by physical forcing (i.e., storms, nutrient availability, insolation, water transparency, water column temperature, etc). This is an excellent method to study the coupling of biological and physical processes.

This is a fundamental oceanic, if not Earth process. The product will require Level-3 MODIS-T and MODIS-N products which are Earth located on a common grid, performed either at EOSDIS or at the TMCF. In either case, the TMCF will have to ingest large data quantities of MODIS and non-MODIS products and data, and equipment provisions must be made available.

2.2.7 *Gelbstoff Concentrations*

2.2.8 *Detritus Concentrations (Dr. Kendall Carder)*

This series of R&D products by Dr. Carder will derive from the same set of algorithms, and so are considered jointly here. Dr. Carder proposes to develop algorithms to identify and quantify these concentrations. Development of the algorithms requires water-leaving radiances (Level 2) from MODIS-T from EOSDIS, preferably on 9-track tape or laser disk, so a direct connection is not required. Development of these algorithms also requires substantial in-situ study, using state-of-the-art optical sensors. From these in-situ observations and the MODIS-T observations, algorithms for these products are proposed to be developed. After development, the algorithms will be sent to EOSDIS to be used in the post-launch processing scenario. Ongoing validation procedures will occur post-launch for four years.

2.2.9 *Sea Surface Temperature Quality Assessment Field (Drs. Otis Brown, Ian Barton)*

Dr. Barton proposes to compare sea surface temperatures (SSTs) from MODIS with those of other satellite sensors (Along-Track Scanning Radiometer (ATSR) on ERS-1, AVHRR, AMRIR), as well as in-situ ship observations to evaluate the reliability of MODIS-N SSTs, and produce an assessment of these SSTs on a global basis. He will thus require calibrated (Level 1B) radiances from

EOSDIS for MODIS-N as well as other satellites, which will be merged at the TMCF. The Quality Assessment Field would then be routed through EOSDIS for use by other investigators. A network link to the CDHF is thus required.

Dr. Brown also proposes to compare MODIS-N SST observations to those from the AVHRR and from ships of opportunity and drifting buoys. However, he requires Level 0 data in addition to Level 1B to perform his quality assessments. He also requires GTS access for drifting buoy SST observations, a pathway for receiving ship of opportunity data, and a communication link to EOSDIS. The Quality Assessment Field generated from the above data sources will consist of the SST and a coded data quality flag and will then be routed through EOSDIS for use by other investigators.

2.2.10 *Objectively Analyzed Sea Surface Temperatures (Dr. Otis Brown)*

Objectively analyzed SST are generated by MODIS-N SST observations coupled with ship and drifting buoy data, coupled with optimal assimilation codes. Consequently, the same data links as for.the SST Quality Assessment Field are required, as are GTS links for in-situ data.

2.3 **Non-Team Member Access**

Once MODIS is launched and the core data products validated, the data will be available for acquisition and use by members of the general scientific community. One may logically anticipate a major demand for, and use of, these data and improvements in algorithms by these non-MODIS Science Team Members. Possible substantial growth of products, methods, algorithms, etc. by these non-Science Team Members may be expected and must be accounted for and facilitated.

We assume that the Science Team Members will play a large role in determining which new data product algorithms enter the MODIS post-launch data processing scheme. If there are conflicting algorithms, i.e., different algorithms to produce the same product, EOSDIS could maintain a library of alternate algorithms to be made available for users' choice.

3 POST-LAUNCH TERRESTRIAL SCIENCES DATA PROCESSING SCENARIO

3.1 Standard Data Products

The algorithms for production of core data products, developed, tested and refined at the TMCFs, will be implemented at the CDHF. It will be the responsibility of the Science Team Members in the TMCF to verify that these core data products are valid, particularly early in the post-launch time frame. These core data products will then be made available to the scientific community through EOSDIS.

3.2 Research and Development Products at TMCFs

The TMCFs will be involved with researching and developing new data products, most probably from Level-1 and Level-2 (atmospherically corrected) data. Any new products developed at the TMCFs, or by other scientists, will probably have to pass a peer review process before being released for production at EOSDIS as new standard data products. This scenario excludes those studies requiring near-real-time data which will require high-priority processing to meet data delivery timeliness requirements.

A simple overview of one aspect of the post-launch processing scenario between a TMCF and the CDHF is depicted in Figure 1; its components are discussed below. It is anticipated that the processing system would be tested with simulated data prior to launch and that the first task in post-launch would be to verify the accuracy of products produced with actual MODIS data.

3.3 Algorithms

The MODIS Science Team Members will develop algorithms to be used for the production of selected core data products at the CDHF. It is anticipated that performance and output of most of these algorithms will have been verified between the TMCF and CDHF prior to launch, and be on line at the time of launch. However, they potentially will be revised during the post-launch period depending upon their performance with actual MODIS data. Provision should be made for revising these algorithms, if justified, shortly after MODIS begins operation, so that accurate core data products can be produced as

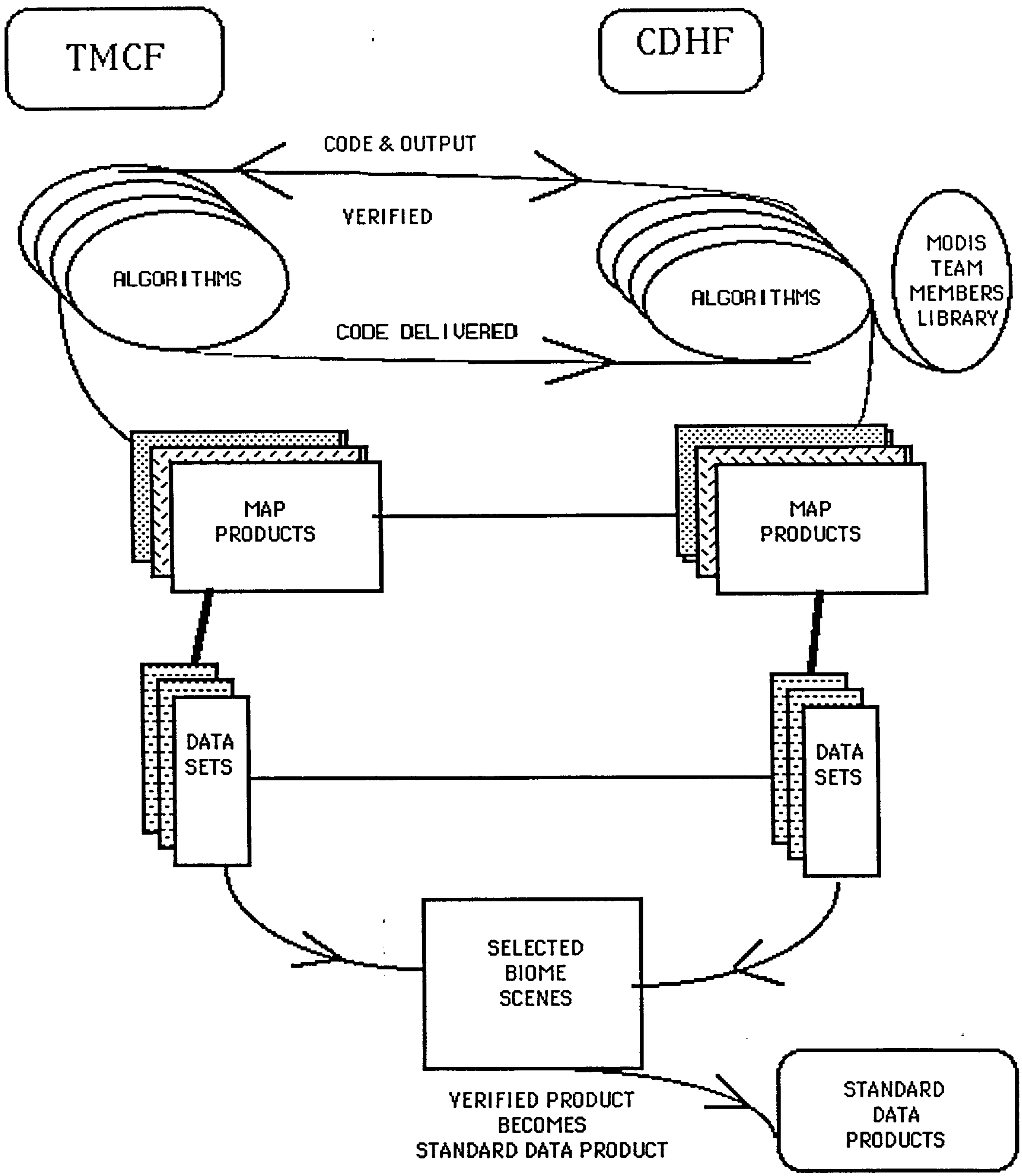

Figure 1. Overview of One Aspect of the Post-Launch Processing Scenario Between the TMCF and the CDHF.

soon after launch as possible. Also, since several team members have proposed to work on similar research problems, e.g. vegetation indices, it is proposed that there be a MODIS team member algorithm library maintained at the CDHF, or within EOSDIS, to facilitate the scientific exchange of ideas among the team members.

3.4 Map Products (Core Data Products)

It appears that most core data products will be in map form (Level-3) and are labeled as Map Products in Figure 1. Also, calling the data products "maps" gives a spatial and temporal emphasis to the core data products. Ancillary data sets, if required for the production of map products, will need to be generated and linked to the map product. MODIS team members will have the responsibility of assessing the accuracy or validity of the first map products, produced with actual MODIS data, using the algorithms they developed, prior to EOSDIS distributing the products on a routine basis.

As a suggestion, these products may be verified on selected biome scenes. Team members have proposed to do research over different biomes; thus, a selection of biome scenes selected for the purpose of testing and evaluating the performance of different algorithms would facilitate scientific exchange among the team members, and other scientists in the future.

Listed below are potential R&D data products identified from team member proposals and previous MODIS documents. Most products have been specifically identified and are preceded by a unique MODIS Data Product and Algorithm code number (Definition of code numbers given in Appendix A). In some cases, only a general product name is given as there may be many variations of the product possible. Also, some products have been proposed by more than a single team member so some product redundancy may appear in the listing.

At present this is assumed to be a reasonably complete list of potential R&D products to be developed in the post-launch phases of MODIS by team members. Some or many of these products may eventually become accepted as core land data products produced at the CDHF, but for the present are expected to be researched and developed independently at the TMCFs. Also these R&D products should be considered as dynamic and subject to change as knowledge advances.

From the discussions presented in team member proposals it is apparent that many of the proposed R&D products depend on, or may be enhanced by, the development of atmospheric correction algorithms for producing land leaving radiances to be used in those proposed R&D products. Atmospheric correction algorithms have been identified as a component of the "utility" algorithms for processing MODIS data and may also be considered as R&D products.

Surface Radiation Budget Components

AL2AS0072	Production mode spectral surface radiances ($Wm^{-2}\mu m^{-1}$)
LL2VS0032	Daily global bidirectional radiances (land leaving radiances ($W\ m^{-2}\ \mu m^{-1}$)
LL3VS0033	Weekly global bidirectional radiance maps ($Wm^{-2}sr^{-1}\ \mu m^{-1}$)
LL3VS0004	Daily Surface reflected solar radiation maps ($Wm^{-2}\ \mu m^{-1}$)
LL3VS0003	Daily surface downward solar radiation maps ($Wm^{-2}\ \mu m^{-1}$)
LL3VS0002	Daily surface upwelling thermal radiation maps ($Wm^{-2}\ \mu m^{-1}$)
LL3VS0001	Daily surface downwelling thermal radiation maps ($Wm^{-2}\ \mu m^{-1}$)
LL3VS0005	Daily surface albedo maps %
LL3AS0067	IPAR (incident absorbed photosynthetically active radiation)
LL2SR0013	Level 2 absorbed photosynthetically active radiation
LL3SR0014	Daily absorbed photosynthetically active radiation maps
LL3VV0042	Polarized part of ground radiance ($W\ m^{-2}\ \mu m^{-1}$)
LL3YK0057	Surface spectral reflectance (6 bands)
LL3AH0074	Soil surface spectral map

Land Surface Temperature

LL2ZW0030	Level 2 MODIS-N land surface temperature °C
LL2SR0028	Level 2 MODIS-N land surface temperature °C
LL3ZW0031	MODIS-N land surface temperature maps °C
LL3SR0029	MODIS-N land surface temperature maps °C
LL3SR0058	MODIS-N weekly maximum NDVI land surface temperature maps °C
LL3SR0059	MODIS-N Weekly average nighttime land surface temperature maps °C
LL3SR0059	MODIS-N Weekly average nighttime land surface temperature maps for cloud free conditions °C

LL3YK0060	Land surface temperature °C for identification of fires (fire products e.g. distribution, extent, etc.)
AL3JS0068	Land and sea surface temperatures °C
LL3ZW0073	MODIS-N land surface spectral emissivity maps

Surface Snow and Ice Cover

LL2VS0034	Level 2 daily global snowcover
LL2SR0037	Level 2 daily North American snowcover
LL3SR0038	Daily north american snowcover maps
LL3VS0035	Daily global snowcover maps
LL3SR0039	Weekly north american snowcover maps
LL3VS0036	Weekly global snowcover maps

Vegetation Indices

LL2CJ0069	Level 2 MODIS-N NDVI
LL3CJ0008	Daily MODIS-N NDVI maps
LL3SR0006	Daily MODIS-T/N NDVI maps
LL3SR0007	Weekly MODIS-T/N NDVI maps
LL3**0009	Greenest MODIS-T/N NDVI maps
LL3AH0063	Weekly soil adjusted vegetation index (SAVI)
LL3CJ0017	Global phenology
LL3VV0043	Polarized vegetation indices
LL4CJ0018	Multi-resolution global phenology
LL4CJ0019	Global growing season length and its interannual variability

Land Cover Maps

LL3AH0075	Vegetation community types
LL3AH0026	Generalized vegetation distribution maps (biome types and soil types)
LL4SR0021	North American biome types
LL4JM0054	Low resolution land cover types maps (biome types)
LL4JM0048	Geolocated spectral signatures
LL4JM0049	Geolocated surface types classed by spectral behavior
LL4JM0050	Geolocated surface types classed by spectral BRDFs
LL4JM0051	Geolocated surface types classed by agro-physical parameters

LL4JM0052	Geolocated surface types classed by un-vegetated mineral composition
LL4JM0053	Geolocated surface types classed by un-vegetated rock type

Other Biological Data Products

LL3SR0012	Daily net photosynthesis maps (t C ha^{-1} day^{-1})
LL3SR0024	Daily evapotranspiration maps (mm H_2O ha^{-1} day^{-1})
LL3SR0025	Weekly North American evapotranspiration maps (mm H_2O ha^{-1} $week^{-1}$)
LL4SR0040	Weekly North American water stress maps
LL3SR0015	Weekly North American net photosynthesis maps (t C ha^{-1} $week^{-1}$)
LL4SR0016	North American annual net primary production map (t C ha^{-1} $week^{-1}$)
LL4JM0020	Canopy leaf area index

3-dimensional vegetation geometry and its change

LL3AH0076	Vegetation density map
LL3AH0077	Soil hydrothermal map
LL3AS0070	Canopy structure
LL3AS0071	Green biomass
LL3AS0072	Woody biomass

Spectral bidirectional reflectance distribution function (SBRDF)

Advanced Studies

Land cover change maps - unspecified

LL4DT0044	Vegetation development study
LL4CJ0023	Land-degradation sub-Saharan Africa
LL4YK0045	Biomass burning studies (%/unit area/unit time)
LL4AH0027	Organic carbon pools
LL4CJ0022	Grassland biome annual production estimates
LL4VS0064	Models of surface water budget including snow/water dynamics
LL4JM0055	MODIS digital elevation models (DEM)
LL4JM0056	MODIS surface roughness maps

Supporting Algorithms

Atmospheric correction algorithms

LLAAS0010	Inversion algorithms for canopy structure
LLAAS0011	Algorithms for estimation of green biomass, woody biomass, IPAR, etc.
LLAVV0041	Atmospheric polarization correction algorithms
LLAVV0065	Photosynthetic capacity algorithm
LLAJM0046	Automate Image Understanding System (ATIUS)
LLAJM0047	Monte Carlo Ray Tracing (MCRT) algorithm
LLAAH0066	Vegetation-detritus-soil spectral decomposition algorithm
LL3ZW0073	Mixed pixel algorithm

4 POST-LAUNCH ATMOSPHERIC SCIENCES DATA PROCESSING SCENARIO

The MODIS team members proposed some data products in the atmospheric sciences which were not identified in subsequent science team meetings as core data products. Interdisciplinary investigators and other scientists may have additional data products which they would like to see become standard MODIS data products. The next two sections tabulate these products. This is followed by a discussion of how the processing and storage may be effected by the new data products. Questions about spatially and temporally limited data products are then raised. This is followed by an estimate of the growth rate in processing for the atmospheric sciences.

4.1 Candidate Post-Launch Data Products By Team Members

Some atmospheric data products were proposed by team members, but did not become core data products. These R&D studies have the potential of becoming new standard data products sometime after launch. A list of some of these candidate future core data products is:

a. Aerosol transport processes (Dr. Kaufman)
b. Aerosol effects on cloud albedo (Dr. Kaufman)
c. Longwave cloud forcing studies (Dr. Susskind)
d. Precipitation studies (Dr. Susskind)
e. Cloud free albedo (Dr. Susskind)
f. Spectral bidirectional reflectance distribution using MODIS-T in stare mode (Drs. Muller, Salomonson, and Strahler)
g. Atmospherically corrected imagery (Dr. Kaufman)
h. Total column precipitable water within clouds (Dr. Kaufman)

The MODIS Science Team Members will develop additional R&D and eventually standard data products besides those listed above.

4.2 Candidate Post-Launch Data Products By Other Investigators

Interdisciplinary investigators and perhaps even members of the wider scientific community may wish MODIS to generate data products which are not now proposed as core data products. Until the interdisciplinary proposals fully evolve, the exact nature of the requirements they may impose on the data processing will not be known. Some potential data products which may be proposed by Interdisciplinary Investigators and others are (the list provide examples only and is not exhaustive by any means):

a. Surface spectral radiation budget components
b. Top of the atmosphere spectral radiation components (global?, regional?, temporally limited?)
c. Aerosol height distribution
d. Surface energy balance components such as sensible and latent heat fluxes
e. Teleconnection studies
f. Mesoscale model predictions
g. Global general circulation model studies
h. Wind speeds from cloud observations

4.3 Examples of the Incorporation of New Atmospheric Data Products

Given that a new data product is proposed to become a standard data product and thus generated upon a regular schedule, how does it become incorporated within the CDHF processing? In many cases the new data product is simply an addition to the processing already being done. For example, aerosol transport requires aerosol loading as an input, so a routine which imports this data and generates the aerosol transport results is all that is required. This routine could be added on at the end of the processing. Most of the team member proposed data products (now classified as R&D) fall into this category. However, these new data products may impose new (and not fully defined) requirements for ancillary data (e.g., NMC upper-air wind fields) that may significantly impact the data processing scenario.

The addition of the spectral bidirectional reflectance function (SBDRF) to the processing is more disruptive to the processing. If this is to be developed from MODIS measurements, the MODIS-T must be placed in a stare mode, meaning that other uses and data processing for these data might be terminated for these time periods. In effect, when the SBDRF is calculated a totally different data flow in the CDHF may be used.

The new standard data product might be most difficult to incorporate in the data processing is one which would appear early in the data flow charts (e.g., Level-1 or perhaps Level-2). At present we do not have an example of this type of data product. If such a data product should appear, the driver program for the processing might need to be re-written. If the processing code was written to be modular, the impact of introducing new data products might be minimized.

The other potential impact of a new data product is that it will be sent to the DADS. It will probably not be sent alone, but might be bundled with other data such as location and times and perhaps with other related data products. For example, it may be desirable to have total precipitable water and liquid water in clouds combined into one data product. The I/O routines for this data product would then need to be re-written.

4.4 The Role of Spatially or Temporally Limited Data Products

Standard data products will be generated for all times and regions. It may be worthwhile to consider standard data products or CDHF generated data products which are generated for limited spatial or temporal domains. These data products may require more computational power than the TMCFs can provide, but may be required for only special studies or field experiments. For example, the surface energy balance components may be calculated for a field experiment in Kansas. Or atmospherically corrected imagery may be desired using the best available radiative transfer models. At present, the requirements for this type of processing is unspecified.

4.5 Projected Future Growth of Atmospheric Data Products

There are seventeen core atmospheric data products. Team members have identified another eight data products which could become new standard data products. Hypothetically, interdisciplinary investigators could have an additional eight data products. If all these data products become standard data products during the five-year mission, the standard data products will nearly double from seventeen to thirty three data products. This rough estimate indicates the growth rate will be about 14% per year. Other rough estimates have suggested a 40% per year growth. This rate of growth (which may be different for each discipline depending on the maturity of the algorithms) must be quantified so that a processing system adequate for five years of growth may be specified.

5 PRODUCT INTEGRATION

5.1 Definition of Product Integration

Investigators working independently and responsible for only one or a few final data products may implement algorithms and intermediate procedures similar or identical to those already implemented by other investigators. We shall use the term product integration to mean a series of procedures and data system design features that reduce data system redundancy. System hardware and software requirements may both be affected by the elimination of redundant data system features.

5.2 Levels of Product Integration

The elimination of redundant data system features can occur either at the input parameter or at the data product level.

Certain geometric and physical parameters associated with an observed pixel will be required inputs for a number of product algorithms. These parameters can be computed once and made available for general use as needed in any of the product algorithms. An example of this type of parameter might be the Solar Zenith Angle. The computation of such parameters is relatively straightforward and uncontroversial. The advantage of such precomputation is the elimination of redundant processing and the simplification of software maintenance. This type of integration requires relatively less coordination between team members.

A second type of product integration requires the use of intermediate or final products generated by one team member as inputs for the generation of another product by another team member (not to mention the use of intermediate or final products generated by one EOS instrument team's algorithms as inputs for the generation of another product by another team's algorithms). This type of product integration obviously requires greater cooperation, coordination, and agreement among team members (and instrument teams).

5.3 Goal and Value of Product Integration

One obvious reason to reduce data system redundancy is to reduce hardware resource requirements. Specifically, requirements for processing capability and data storage will both be reduced by the elimination of redundant processing and products. Software initial implementation complexity and maintenance requirements will also be affected. However, since product integration may require considerable coordination and agreement among investigators, it is not obvious that product integration will result in overall simplification of the software implementation process. Indeed, one could perhaps argue that the generation of integrated software may require a higher level of software coordination and support from central EOSDIS facilities than that needed for strictly independent product generation by individual team members.

Some obvious areas where integration may be applied are in cloud identification, atmospheric correction, and Level-2 to Level-3 product conversion (gridding and time and space averaging).

5.4 Product Integration at the TMCF

Since products implemented at the CDHF after platform launch will be developed at the Team Member Compute Facility (TMCF), cost-effective product integration must begin at the TMCF. A number of general data system features would facilitate the development of TMCF code that is readily transportable to the CDHF.

One obvious feature is that the operating system chosen for the TMCF and CDHF should support the development of transportable code between the two systems. At present, no programming language requirement for EOSDIS has been formally stated. FORTRAN and C are often mentioned as potential language standards. UNIX is often mentioned as an operating system that supports both of these languages. Unix is the de facto operating system standard for scientific workstations and may therefore enjoy wide support among MODIS Science Team Members and the larger data user community.

Another requirement is common data formats for all data exchanged between the TMCF and other EOSDIS elements. Once suitable general formats for data are identified, it appears that the implementation and use of these formats at the TMCF will pose few special problems. If code is to be developed at the TMCF and transported to the CDHF, the advantages of standard data formats for use throughout EOSDIS are obvious.

Common software standards also support software transport. It is expected that the Science Data Support Team (SDST) will play an important role in the implementation of software at the CDHF and might serve as the enforcing agency for agreed-upon software standards. Compliance with software standards at the TMCF might be left to the individual discretion of the Team Member, but it is thought that compliance with standards is best built into software during the development process, so that the most cost-effective approach for software to be implemented on the CDHF will be to maintain standards compliance throughout the development process.

Several specific data system features at the TMCF would also support the development of integrated software. One requirement is that common input parameters that will ultimately be available for use at the CDHF should also be available to the investigator working at the TMCF. The investigator can then integrate the use of such precomputed numbers during the code development and validation process. A similar but less well defined requirement exists for software integration at the product level. To take advantage of products available or planned to be available from other team members, a team member must be able to access descriptive information and sample representative results for the product in question. Team members must be willing to work closely and the data system itself must support coordination of efforts, i.e. data transmission capabilities between the cooperating facilities must be adequate to support coordination of efforts.

5.5 Product Integration at the CDHF

Although software development occurs primarily at the TMCF, several system features at the CDHF may also facilitate the development of integrated software.

Peer review of products prior to implementation on the CDHF would provide an obvious opportunity to review the analysis behind a product and identify common or diverging elements in various team member approaches. Common and diverging elements may be flagged for further study, and, depending on the results of such further investigations, unified approaches may evolve that decrease overall system hardware and software maintenance requirements.

The existence of the SDST will also tend to support the development of integrated software. The SDST can serve as a clearing house for previously developed algorithms that may be applicable to the present needs of an investigator working at his TMCF. In such cases, the team member may prefer to use an existing procedure rather than develop, implement, and document an independent approach.

6 POST-LAUNCH DATA CALIBRATION SCENARIO

The visible, near-infrared and thermal-infrared bands of MODIS-N and MODIS-T must be continually monitored in the post-launch period to ensure a high consistency of the Level-1B calibrated radiances. This requires ongoing and continuous calibration procedures (Figure 2). These calibration checks should be stored at the Level-1A processing stage to allow investigators to perform their own checks.

Calibration will be a continuing task in the production of Level-1 data products. Calibration occurs first in the processing scenario as a task shared by the TMCF, Calibration Support Team (CST), and the CDHF. Calibrated data will then be delivered to the TMCFs.

Some Science Team Members have expressed a desire to apply their own calibrations to the Level-0 data. They would then produce their own Level-1B products, which implies that these products will differ from those produced by the CDHF.[3] The MODIS calibration software can be divided into several major categories:

- Ground Support Software: Provides analyses of the components and assembled instruments prior to launch.

- Activation Period Software: Provides initial calibration information on the instruments during the first few months of operation.

- Routine Calibration Software: Provides the calibration coefficients used for routine data processing.

- Verification Software: Provides comparisons of the MODIS radiances to radiances from other platform instruments, other satellites, or in-situ observations.

---

[3]Dr. John Parslow in Australia may require direct data downloading to have timely access to MODIS observations. Since the data are directly downloaded, he may apply his own calibrations. His Science Team Member Computing Facility (TMCF) may require additional computational capacity to permit them to generate daily regional Australian products from Level 0 data.

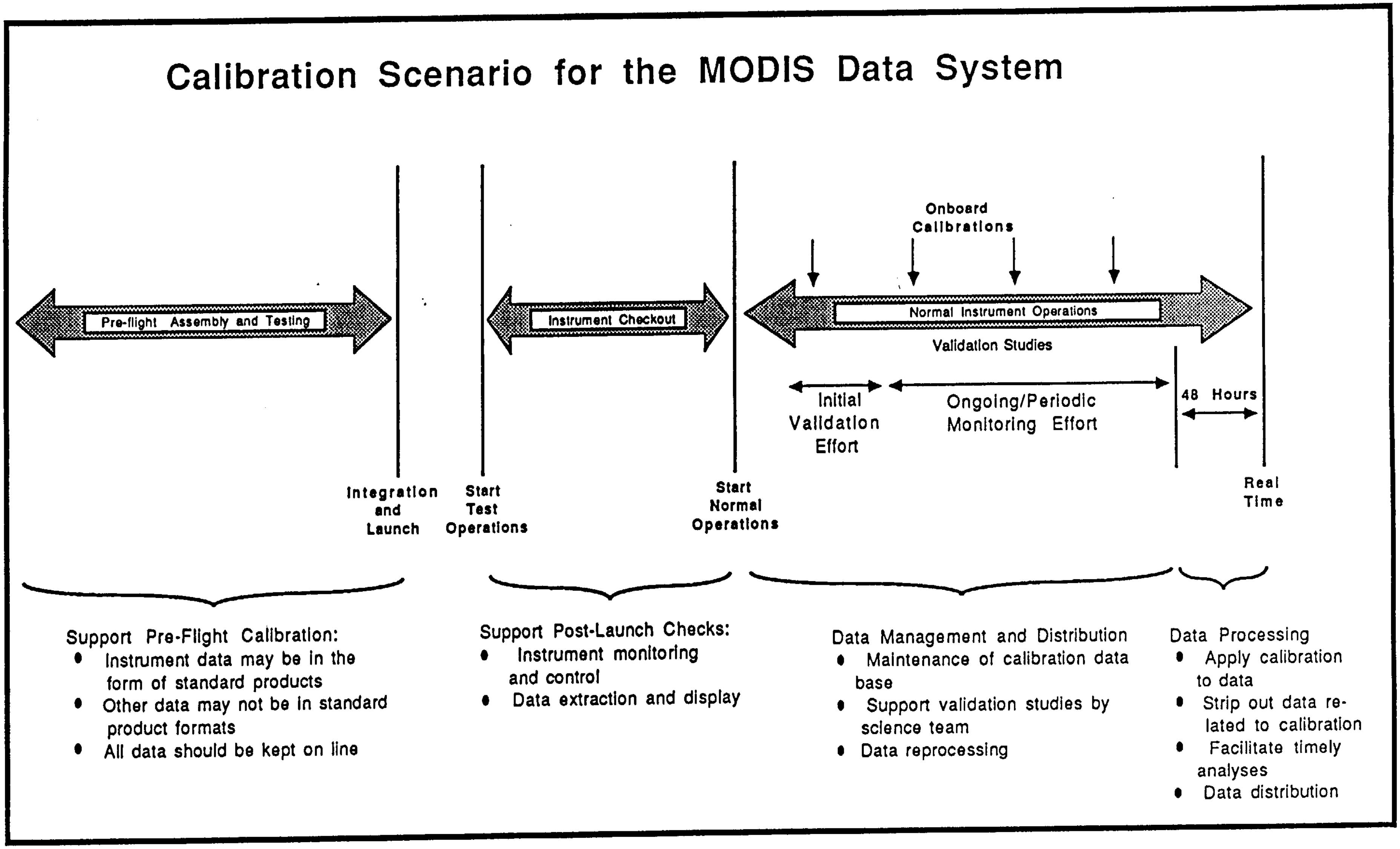

Figure 2. Calibration Scenario for the MODIS Data System

- Diagnostic Software: Provides analyses of the instrument or components both prior to and after launch.

With sufficient fore-thought, most of the software above will be developed prior to launch. If an unexpected problem arises, it will probably be handled by modifications of existing software or the development of new diagnostic software.

Some characteristics of the post-launch calibration and verification software will be discussed in the next few sections.

6.1 Activation Period Calibration Algorithms

The following tests will be done in the activation period:

- All calibration modes tested and intercompared
- All calibration sources checked
- Spectral calibrator checked
- Calibration algorithms checked
- Initial Level-1B data generated
- Initial measurements of pre-selected Earth targets performed
- Solar diffuser plate scanned during a satellite yaw maneuver

All but the last of these calibration procedures will be performed routinely during the operational phase of the instrument. It is likely that the calibration software will have options within it that can handle the tests above.

6.2 Routine Calibration Algorithms

The following normal operation modes for both MODIS-N and MODIS-T are expected:

- Day or night mode (Earth-viewed)
- Solar calibration mode (diffuser plate deployed and viewed)
- Lunar calibration mode (full moon seen through aperture)
- Spectral calibration mode (spectral calibrator on)

Each of these modes will require special software utilities to extract and interpret the data streams. The following ancillary data is associated with the calibration modes:

- Lamp(s) output, voltages, and currents
- Lamp detector outputs and temperatures
- Active cavity radiometer outputs and temperatures when solar diffuser plate is deployed
- Spectral calibrator output and temperatures

It is not clear if these outputs will always be downlinked or will be downlinked only when MODIS-N is in one of its calibration modes. In either case, special software to extract and analyze the ancillary data will be required.

In addition to the above, MODIS-T may have several modes of operation unique to itself. These are:

- Tilt modes (any angle from 0 to $\pm$ 50 degrees)
- Stare mode (one location on Earth is viewed continuously as the satellite passes over for SBDRF studies)
- Sun glint avoidance mode (the tilt changes to avoid the sun-glint and missing data is filled with MODIS-N data)
- Stereo mode (for stereo images of single Earth region)

In each of these modes, the normal data stream is probably transmitted (except the sun-glint avoidance mode), so software will be required to extract the data and properly analyze it.

6.3 Verification Algorithms

For verification studies, the normal data is transmitted, calibrated, and stored. When a verification study is performed, software to extract the data, given the locations and times desired, is required. The extracted data can then be compared to the data from other satellites by an interactive image processing station which allows overlays, spatial filtering, spectral manipulations, and so forth. Alternatively, if a type of comparison is performed frequently, the process can be automated and done in a batch mode. Therefore, two categories of verification software can be anticipated:

- Interactive image processing
- Batch radiance comparisons

The following types of software must be available:

- Mapping software to overlay, rectify, and filter radiances from other satellite instruments such as GOES, AVHRR, Landsat, and HIRIS.
- Mapping software to convert point surface measurements to a MODIS compatible coordinate system
- Radiative transfer codes (analytic, Monte Carlo, etc.) to convert measured surface leaving radiances from field experiments to satellite level radiances
- Modulation transfer function codes
- De-striping software (possibly)
- Day-night difference software
- Stereo cloud height versus thermal cloud height comparison software
- Sun-glint and desert target software

In the post-launch period, it is anticipated that new satellites will be launched which will make radiances measurements for which a comparison to MODIS radiances will be desirable. These could include various European and Japanese platforms, for example. Software to import these radiances and put them in a standard MODIS format may require development. The extent of this developmental work is as yet not determined.

6.4 Projected Future Growth of Calibration Data Products

There is no estimate of the growth in calibration data products after launch. The development of new software for calibration and verification studies is planned to be on-going throughout the mission.

7 POST-LAUNCH DATA REPROCESSING SCENARIO

This section presents our current understanding of reprocessing. The reasons for doing reprocessing (calibration changes, algorithm changes, ancillary data changes, and even retrospective reprocessing due to the introduction of new data product algorithms) are presented. A number of issues and questions on the data management required for and associated with reprocessing are discussed. The final subsection considers the implications of reprocessing on the capacity required for the CDHF.

7.1 Calibration Changes

If and when it is necessary to change or correct the calibration of the MODIS data, extensive reprocessing will be required. There are two ways in which the calibration can be changed. The simpler method is to use different calibration coefficients. The more complex method would use a different calibration procedure with a different number of coefficients and/or different equations for calibration. The data management and processing capacity issues will be discussed in a separate section.

7.2 New Coefficients

The simple type of recalibration would be done by using revised calibration coefficients. In the scenarios which we have developed, it has been assumed that the calibration coefficients are stored in the header of the Level-1A data cube. If only the values of the coefficients change, reprocessing would require that the entries in the header be revised and the calibration could be redone using the same procedure as used in the initial processing.

This type of reprocessing is expected to proceed as follows. (See data management section for additional discussion.) The Level-1A data will be recovered from the DADS or permanent archive. The revised calibration coefficients will replace the old values in the header. The reprocessed Level-1A data could then be sent through the standard processing which would regenerate all of the products which were produced during the prior processing. (This scenario assumes that the standard processing software has not changed since the previous processing.) The Level-1, 2, 3, and 4 data products generated would replace (or exist along-side) the previously stored values. The reprocessed data would have a volume and format identical to the previous version.

Any archived data might or might not be overwritten, depending on the reprocessing policies of EOSDIS and the MODIS Science Team.

It is likely that a change in calibration coefficients would not affect all of the channels. This would imply that not all of the data would need to be recalibrated. There could be some higher level products which would not be changed as a result of recalibration. It is our view that the data management will be so complicated that it might be recommended that all products be recalculated. However, it would be possible to determine which products needed to be recalculated.

In summary, when calibration coefficients are changed the header in the Level-1A data will be changed and all of the higher level products calculated. The reprocessed data would replace the previous stored data and data products.

7.3 New Calibration Procedure

It is possible that reprocessing of data will be required due a change in the calibration procedure. This would involve the implementation of new calibration equations. A revised set of calibration coefficients would be used and the number of coefficients could change.

The Level-1A data would be recovered as the starting point for reprocessing. If the number of coefficients changes, it would be necessary to modify or replace the existing header. The algorithm used to recalibrate the data will be different from the that used in the initial processing. The processing control software may have been changed to properly read the new header format.

The reprocessed data will again replace the previous version of the data. However, if the header has been changed significantly, the data may no longer fit into the area where it was stored. Any change in the volume of the data during reprocessing may require significant changes in how the data are stored in the DADS. This is an important issue. Certain types of reprocessing may require that the data archive be reformatted.

7.4 Algorithm Changes

Algorithm changes will be the second major reason for reprocessing. There are three types of algorithm changes presented in this discussion: error correction, algorithm updates, and new algorithms.

7.5 Correction of Algorithm Errors

With the volume of computer code that will be required to generate the MODIS data products, it is certain that there will be some undetected errors in the code as implemented on the CDHF. The correction of these errors might require that the data be reprocessed (or documented).

The reprocessing will begin with the data that are required as input to the corrected algorithm. As an example, the input to a weekly composite product would be seven days of Level-2 data. The standard processing sequence would be started at the point of the corrected algorithm. The corrected product and all the products calculated using that product as input will be generated. The reprocessed products might replace the previous version in the archive.

This type of reprocessing will require that it be possible to start the standard processing sequence at the position of the corrected algorithm. It would also require the ability to either overwrite selected products in the archive or else provide effectively new products to the archive. This implies a set of process control software that is substantially different than that used in the standard processing.

7.6 Updated Algorithms

There will be changes made to algorithms. For example, the resolution of a weekly composite map might be changed from 8 to 4 kilometers. This would double the volume of the particular Level-3 product. An algorithm might also be changed to use a different source of ancillary data.

This type of reprocessing would require more significant changes in the data processing control software. In the second example, reprocessing would require that ancillary data be used which were not stored in the header of the data cube. The control software would need to identify the location of, and obtain, the required ancillary data. (Reprocessing could not be done if the ancillary data were not available.)

In the first example above, the data volume of the particular Level-3 product would be increased. The reprocessed data would not fit in the storage location of the original data. (See Section 7.9 on data management for further discussion of this issue.)

7.7 New Algorithms

There will be additional MODIS algorithms developed and implemented after launch. A majority of the Team Members' proposed data products are not included among the core data products. Furthermore, it is expected that some of these products will be much more complicated than the core data products.

New algorithms will be developed and installed in the standard data processing. It is likely that some or all of the previously collected data will be reprocessed to generate this newly installed product. The data management problem will be particularly difficult in this situation). The new product will not replace a previously archived product. Retrospective processing will often be required to obtain multi-year time series. New input ancillary data sets may be required, perhaps from new data sources.

7.8 Ancillary Data Changes

Ancillary data will be used in generating most of the geophysical data products. A primary source of ancillary data will be other satellites and other EOS instruments. The values of the ancillary data may change due to reprocessing done on those experiments. When there is a significant change in the ancillary data, it will be necessary to reprocess MODIS data to regenerate products with the corrected values of non-MODIS input data.

This reprocessing will be done by retrieving the appropriate level of data from the archive, replacing the ancillary data with the corrected values, and regenerating the affected product and those products that follow. This type of reprocessing will require that the standard processing be started at the appropriate point in the processing cycle. Process control software similar to that required for algorithm correction will be required.

7.9 Data Management Issues

There are a number of unresolved issues on how the data will be managed during reprocessing. These issues are presented in this section beginning with a brief overview of how the data will be managed in standard processing.

The standard processing should be fully automated and controlled by a process management function in the CDHF. The raw MODIS data will be received, calibrated, and combined with any needed ancillary data. The assumption has been made that any non-MODIS data needed to generate standard products could be written to and retained in the header of the MODIS data and/or data products. (Or otherwise available upon demand at the time of processing.) The process control software will drive the proper sequencing of the data processing and ensure that all of the inputs to a given algorithm are available.

It is assumed that the raw data, all of the derived data products, and all of the required ancillary data products will be archived together. Metadata, browse data, and catalog data will be generated and sent to the IMC.

Reprocessing will require different process control software. The degree of difference will be determined by the exact type of reprocessing to be done.

Here is a specific yet hypothetical example. In standard processing, the atmospheric correction algorithm used as part of ocean data product generation requires surface pressure with an accuracy of 1 mb. The pressure in this hypothetical case is obtained at close to this accuracy by objective analysis at the National Meteorological Center of global station observations of surface pressure. The standard processing control software waits until the pressure is available before proceeding with the atmospheric correction.

Four years after the launch of MODIS, an improved atmospheric correction algorithm with higher accuracy (possibly with surface pressures obtained by analysis of AIRS/AMSU data) is perfected and the decision is made to reprocess all the ocean products. Only the ocean data products will be reprocessed. The processing control software will be required to obtain the data from the DADS and permanent archive, start the processing at the point where atmospheric correction is done, and recalculate only the ocean products.

The needed AIRS/AMSU data may be stored as ancillary data in the MODIS data record. If not, the ancillary data must be recovered from the DADS and/or permanent archive where it is stored. This will require that the process control software interact with the IMC to locate and request the desired data. The control software will also generate only ocean data products.

This example makes it clear that the process control software for reprocessing will be fundamentally different from the standard processing control software. It is possible that each reprocessing task will require modifications to the reprocessing control software.

In the above discussion, the reprocessed data may or may not replace the previous version of the products. It has not been decided which procedure will be followed. However, there are reasons why replacement should or should not be done. First, reprocessing will be done to correct errors in either the calibration or product generation. There does not appear to be any reason to retain products that were wrong. However, studies of global change involve the analysis of multi-year data records. A slightly biased 12-year time series is of considerably greater interest then an unbiased six-month partially reprocessed parameter. At a minimum, the earlier data set might be preserved in its entirety until the reprocessing is completed (and then purged). In the interim, both data sets, suitably documented, could be made available. If reprocessing proceeds at twice the rate of data acquisition (five years to reprocess ten years of data), then significant time periods are involved.

Another reason for replacing data on reprocessing is to reduce the volume of data to be stored. The system is being designed to allow for two reprocessings of the data. If all of the results are retained, the volume of the data archive would be tripled.

It may not always be possible to simply replace old data with reprocessed data. If the volume of the reprocessed product is larger than the old version, the new data will not fit in the place of the old data. This would require either than the archive be reformatted or that techniques be developed to manage data when the set of products is not stored together.

When the standard products are generated, the raw data and all of the products generated from that data might be archived in a contiguous storage area. This would be an efficient storage method and will greatly simplify the data management problem (e.g., it would be possible to physically locate any data product in the archive by some simple criteria such as observation time). If products from a single set of observations are in a variety of locations, the data management becomes much more complex. It is likely that the data management issues are so difficult that the archive will be reformatted when necessary to keep data sets together.

Metadata will be generated during all reprocessings and in some cases browse data will be generated. This data will be used by the IMC and is assumed to replace the previous versions. The record of all processing done on any data will be retained as part of the data history.

There are a couple of issues on data distribution and reprocessing. There will be data in the archive to be reprocessed but for which the reprocessing has not been completed. Will these data be available for distribution? As an example, it may be necessary to reprocess to correct calibration errors. The archived data will be inaccurate at best. It might be reasonable to hold these data until it is reprocessed (i.e., don't send out inaccurate data). On the other hand, the reprocessing could take a long time (one year to recalibrate two years worth of data) and the scientists may not be willing to wait to do their science.

The second issue regards data distribution. The IMC will retain a record of what data have been sent and to which scientists. It will be necessary to at least inform the scientists who have received the data when it is reprocessed and redistribution of the data may be necessary. It is unclear how the distribution of reprocessed data will be done.

The final issue on data management can be referred to as the "processing tree". There will be a complex set of dependencies on the various data products. This will include not only the dependencies among MODIS data products but also the products which MODIS requires from other sources and the MODIS products used by other instruments and/or science teams. The reprocessing of a low level MODIS product will require the recalculation of higher level MODIS products plus all of the non-MODIS products that use the new MODIS products as input. There will also be reprocessing of MODIS products as a result of changes in ancillary products. The interdependence of the various instruments must be given careful attention.

Finally, in a holistic sense, it may be useful to define reprocessing cycles that apply to the MODIS processing as a whole, or even to the EOSDIS as a whole. This would avoid what might otherwise be inefficient data management with many data products each in various stages of reprocessing, each with its own I/O and processing requirements.

In summary, there are a number of issues on the data management required for reprocessing. They are:

a. Process control software
b. Data replacement
c. Archive reformatting
d. Data access during reprocessing
e. Data distribution after reprocessing
f. Processing tree
g. Holistic reprocessing strategy

7.10 Processing Capacity/Requirements

This section contains a discussion of the needed capacity for reprocessing and the requirements on the CDHF needed to accomplish this task. Unfortunately, the topic of reprocessing is not well developed so quantitative estimates cannot be made.

The Level-1 documents require that the CDHF be designed to allow all data to be reprocessed twice. This requirement has been included in the preliminary sizing estimates by multiplying the standard processing estimate by three. This is a very rough estimate. The several types of reprocessing will require varying levels of computing capacity.

Only calibration changes will require that most of the initial calculations be done during reprocessing. Even in the case of calibration changes, some of the calculations (e.g., Earth location or cloud detection) would not be repeated. The reprocessing of higher-level products (e.g., primary production) will start well along the processing sequence. There may be much less processing involved in recalculating a Level-3 or -4 product then was required in the initial calculation. The factor of three used in sizing the CDHF may be an overestimate.

The capacity required for reprocessing is a major driver on the system design. An accurate estimate of reprocessing is necessary to generate a defensible estimate of the required CDHF capacity. The reprocessing concepts must be developed and scenarios generated.

It will not be possible to do reprocessing without "good" justification. There must be a formal mechanism, such as peer review, for submitting and approving reprocessing requests. Reprocessing requests will only be done with adequate scientific justification. The effects of any reprocessing requests on the CDHF must be considered. Even the simplest reprocessing may have significant impact on the performance of the CDHF. The prioritization of reprocessing requests can only be done for the EOS project as a whole.

The CDHF resources used for reprocessing will be in an operating environment separated from the resources used in the standard data processing. The data flows will be different, different process control software will be used, and different types of processing may be done. The operations concepts and scenarios, developed for the standard product generation, do not apply to reprocessing without some modification.

There are many questions and issues concerning reprocessing that need to be answered. Obtaining these answers will be difficult since a reliable estimate of changes in the calibration and algorithms is required. A defensible estimate of the capacity required for reprocessing is absolutely necessary if the CDHF is to be properly sized.

ACKNOWLEDGEMENTS

Over the course of three formal MODIS Science Team meetings, much of the guidance in shaping these requirements has been received from the team members, whom we wish to acknowledge here: Mark Abbott, William Barnes, Ian Barton, Otis Brown, Kendall Carder, Dennis Clark, Wayne Esaias, Robert Evans, Howard Gordon, Frank Hoge, Alfredo Huete, Christopher Justice, Yoram Kaufman, Michael King, Paul Menzel, Jan-Peter Muller, John Parslow, Steven Running, Philip Slater, Alan Strahler, Joel Susskind, Didier Tanre, Vern Vanderbilt, and Zhengming Wan. Other members of the MODIS Data Study Team, Lee Kyle, John Barker, and Jim Ormsby, also played an active role in collecting the information presented in this article. The efforts of Brenda Vallette in preparing this manuscript are also acknowledged.

APPENDIX A

MODIS DATA PRODUCT AND ALGORITHM CODE NUMBER DEFINITION

MODIS Data Product and Algorithm code number definition is as follows:

Product Category

C	Calibration
O	Ocean
L	Land
A	Atmosphere

Product Level

L1	Level-1A/B
L2	Level-2
L3	Level-3
L4	Level-4
LC	Calibration
LA	Algorithm
LS	Simulated Data

Originating Team Member

MA	Mark R. Abbot
IB	Ian Barton
OB	Otis B. Brown
KC	Kendall L. Carder
DC	Dennis K. Clark
WE	Wayne Esaias
RE	Robert H. Evans
HG	Howard R. Gordon
FH	Frank E. Hoge
AH	Alfredo R. Huete
CJ	Christopher O. Justice
YK	Yoram J. Kaufman
MK	Michael D. King
WM	W. Paul Menzel
JM	Jan-Peter Muller
JP	John Parslow

SR Steven W. Running

PS Philip N. Slater

AS Alan H. Strahler

JS Joel Susskind

VS Vincent V. Salomonson

DT Dedier Tanré

VV Vern Vanderbilt

ZW Zhengming Wan

Sequential Identification Number

00001-9999